1 [Counsel on signature page]

2

3

4

5

6

7

8   **UNITED STATES DISTRICT COURT**
                    NORTHERN DISTRICT OF CALIFORNIA
9
                          SAN JOSE DIVISION
10

11   Master File Case No. 5:23-cv-03440-EKL
                                            Consolidated with Case No. 5:24-cv-02531-EKL
12
                                            SECOND AMENDED CONSOLIDATED
     In Re Google Generative AI Copyright Litigation   **CLASS ACTION COMPLAINT**
13

14   Judge: Hon. Eumi K. Lee

15
                                            Second Amended Consolidated Complaint Filed:
16   ~~December 20, 2024~~September 25, 2025

17   **DEMAND FOR JURY TRIAL**

18

19

20

21

22

23

24

25

26

27

28

TABLE OF CONTENTS

Plaintiffs Steve Almond, Sarah Andersen, Burl Barer, Jessica Fink, Kirsten Hubbard, Hope Larson, Mike Lemos, Jill Leovy, Connie McLennan, and Jingna Zhang, on behalf of themselves and all others similarly situated (collectively, "Plaintiffs"), bring this ~~consolidated class action complaint~~Second Amended Consolidated Class Action Complaint ("Consolidated Complaint") against ~~Defendants~~Defendant Google LLC ~~and Alphabet Inc. (together, "Defendants" or "~~("Google").

I.   INTRODUCTION

1.   Google willfully infringed millions of registered copyrighted works to create its commercial artificial-intelligence ("AI") business. Plaintiffs are visual artists and authors who have copyright registrations for certain works (including the "Plaintiff Works[1]" cited herein). Google willfully copied the Plaintiff Works without authorization to build and train its generative artificial intelligence models ("Generative AI Models"), including Gemini[2] and Imagen,[3] thereby violating 17 U.S.C. § 501.

2.   The scale of Google's copyright infringement is unprecedented. To build its Generative AI Models, Google copied millions of copyrighted works and permanently embedded copies of these works within these models. For example, Google's training dataset for its Bard model comprised 1.56 trillion words.[4] Similarly, LAION-5B—a likely training dataset for Imagen—comprises 5.8 billion image-caption pairs.[5] The training of a Generative AI Model necessarily involves making multiple unauthorized copies of each work. For each work used in training, first a copy is made during data

---

[1] Section III alleges infringed works for each named plaintiff. Collectively, they shall be referred to as Plaintiffs Works.

[2] Gemini is a Google Generative AI Model that was originally released as a chatbot in March 2023 under the name Bard. Gemini is now a multimodal model family that includes Gemini 1.0, Gemini 1.0 Pro, Gemini 1.0 Ultra, Gemini 1.0 Nano, Gemini 1.5 Pro, Gemini 1.5 Flash, Gemini 1.5 Flash-8B, and Gemini 2.0. Bard and all versions of the Gemini model family are collectively referred to in this Complaint as "Gemini."

[3] Imagen is also a Google AI product, originally released in May 2022, a text-to-image diffusion model family that includes Imagen, Imagen 2, and Imagen 3. All versions of Imagen are collectively referred to in this Complaint as "Imagen."

[4] Romal Thoppilan et al., *LaMDA: Language Models for Dialog Applications* at 2, GOOGLE (Jan. 20, 2022), https://arxiv.org/pdf/2201.08239.

[5] Christoph Schuhmann, *LAION-5b: An Open Large-Scale Dataset for Training Next Generation Image-Text Models* at 2, LAION (Oct. 16, 2022), https://arxiv.org/abs/2210.08402.

1 collection, then a series of copies are made during model training, and finally a copy is permanently

2 incorporated into the parameters of the Generative AI Model.

3   3.   Google's unauthorized copying of the Plaintiff Works continues to harm Plaintiffs

4 through Google's ongoing monetization of those works. Google is integrating its Generative AI Models

5 into a growing set of Google products, thereby powering them with infringing copies of the Plaintiff

6 Works. Those products include Google Search, Google Cloud, Gmail, Google Docs, Google Ads,

7 Google Slides, Chrome, YouTube, Google Photos, Google Sheets, Google Meet, Google Pixel, Google

8 Maps, Google AI Studio, Google Vids, Google Workspace, and Vertex AI (collectively, the "AI-Powered

9 Products").[6] Google was able to create the AI-Powered Products because of the Generative AI Models it

10 built and trained on Plaintiffs' copyrighted works without Plaintiffs' license or authorization.

11   4.   By embedding the Plaintiff Works into its Generative AI Models, Google has irreversibly

12 entangled the Plaintiff Works with its commercial products. Google has stripped Plaintiffs of their

13 exclusive rights under the Copyright Act to control the copying and distribution of the Plaintiff Works.

14 This unauthorized copying and distribution inflicts immediate and irreparable harm on Plaintiffs—harm

15 that Google compounds daily through its expanding deployment of AI-Powered Products built on

16 infringed Plaintiff Works.

17   5.   The FTC issued a stern warning to the AI industry in May 2023 regarding the industry's

18 sudden sprint to collect as much training data as possible: "Machine learning is no excuse to break the

19 law . . . The data you use to improve your algorithms must be lawfully collected . . . companies would

20 do well to heed this lesson."[7]

21   6.   Rather than heed the FTC's warning and cease its years-long theft of data, Google instead

22 elected to continue copying and downloading the works of writers, artists, and other creators—without

23 notice, consent, or compensation—to build and train its AI Products, including Imagen and Gemini.

24 _____

25 [6] *Gemini*, GOOGLE, https://ai.google/get-started/gemini-ecosystem/#GeminiInOurProducts (last visited December 10, 2024).

26 [7] *Statement of Commissioner Alvaro M. Bedoya Joined by Chair Lina M. Khan and Commissioner
Rebecca Kelly Slaughter*, FED. TRADE COMMISSION (May 31, 2023), https://www.ftc.gov/legal-

27 library/browse/cases-proceedings/public-statements/statement-commissioner-alvaro-m-bedoya-joined-
chair-lina-m-khan-commissioner-rebecca-kelly-slaughter-0.

28

Case No. 5:23-cv-03440-EKL   2
SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

7.   When gathering training data for its Generative AI Models, Google chose to copy data from websites under active prosecution for copyright infringement.[8] For example, C4, a dataset Google admits it used to train Gemini, contains materials from Z-Library, a website which hosted pirated copyrighted material. Z-Library now displays this seizure warning:

8.   Google's unauthorized copying of the Plaintiff Works to train its Generative AI Models has created two rapidly increasing revenue streams for Google. First, Google collects subscription and licensing fees from customers of its Generative AI Models. Second, and more significantly, Google has leveraged its Generative AI Models throughout its broader portfolio of AI-Powered Products, resulting in revenue growth attributable to the infringing technology.

9.   Google has confirmed the scale of these commercial benefits through its own financial disclosures and executive statements. In July 2024, Google and Alphabet CEO Sundar Pichai reported that Google Cloud, one of Google's AI-Powered Products, achieved quarterly revenue of $10.9 billion—

---

[8] Ernesto Van der Sar, *FBI Carries Out Fresh Round of Z-Library Domain Name Seizures*, TORRENT FREAK (May 30, 2024), https://torrentfreak.com/fbi-carries-out-fresh-round-of-z-library-domain-name-seizures-240530/.

a 29% year-over-year increase directly linked to AI integration.[9] Pichai also said that Google Cloud's AI

infrastructure and "generative AI solutions, alone, had year-to-date already generated billions in

revenues and are being used by more than two million developers."[10] While Google's campaign of

massive copyright infringement is already generating the intended financial gains, the creators whose

copyrighted works underpin and sustain the Generative AI Models continue to receive nothing.

10.   In October 2024, Pichai confirmed that Google's AI-Powered Products were "paying off

and driving success" by substantially contributing to record-breaking quarterly revenue of $88.3

billion.[11] This financial performance demonstrates the substantial commercial advantage Google has

derived from its unauthorized copying of the Plaintiff Works and establishes the causal link between

Google's copyright infringement and its monetary gains.

Class Certification Presents the Superior Method of Adjudication

11.   Google's uniform approach to data collection and training for its Generative AI Models

compels class certification.[12] Google relied on automated systems to make unauthorized copies of the

Plaintiff Works and millions of others, creating common questions of law and fact that predominate over

any individual issues. Google did not selectively target or differently handle specific Plaintiff Works—

rather, it applied consistent methods across Plaintiff Works that it copied to build and train its Generative

AI Models.

12.   The economics of copyright litigation make individual suits impracticable for most

copyright holders. Pursuing individual copyright claims would require copyright holders to shoulder

---

[9] Nico Grant, *Alphabet Reports 29% Jump in Profit as A.I. Efforts Begin to Pay Off*, N.Y. TIMES (July 23, 2024), https://www.nytimes.com/2024/07/23/technology/alphabet-google-earnings.html.

[10] Lindsey Wilkinson, *Google Ties Billions in Revenue This Year to Generative AI*, CIO DIVE (July 24, 2024), https://www.ciodive.com/news/google-cloud-revenue-growth-continues-generative-ai/722310/#:~:text=%E2%80%9CYear%2Dto%2Ddate%2C,Pichai%20said%20during%20the%20call.

[11] Nico Grant, *Alphabet Revenue Jumps 15% to $88.3 Billion*, N.Y. TIMES (Oct. 29, 2024), https://www.nytimes.com/2024/10/29/technology/alphabet-google-earnings.html.

[12] Class proceedings promote judicial efficiency by avoiding thousands of duplicative suits addressing identical legal theories and evidence regarding Google's AI training practices. Individual actions would unnecessarily burden the courts with repetitive litigation of the same core issues concerning Google's standardized conduct.

1 substantial litigation costs that would likely exceed potential individual recovery. Class treatment allows

2 copyright holders to pool resources, achieve economies of scale in litigation, and enable meaningful

3 access to justice. Class certification would also advance the Copyright Act's fundamental purpose of

4 protecting copyright holders by providing a viable mechanism to challenge systematic infringement.

5 Without class treatment, most copyright holders would face insurmountable practical barriers to

6 vindicating their rights, effectively immunizing Google's widespread copyright violations from judicial

7 scrutiny.

8   13.   The standardized nature of the infringements, the need for effective enforcement of the

9 U.S. Copyright Act, and the practical barriers to individual suits all point decisively toward class

10 treatment as the superior method for adjudicating these claims.

11   14.   Plaintiffs therefore bring this action to recover statutory and actual damages, Google's

12 profits attributable to its infringing conduct, attorney's fees and court costs, injunctive relief, and all

13 other relief deemed appropriate by the Court on behalf of themselves and all those similarly situated.

14 **II.   JURISDICTION AND VENUE**

15   15.   This Court has subject-matter jurisdiction under 28 U.S.C. § 1331 because this case arises

16 under the Copyright Act (17 U.S.C. § 501). Jurisdiction and venue are proper in this judicial district

17 under 28 U.S.C. § 1391(c)(2) because ~~Defendants are~~Google is headquartered in Mountain View,

18 California and a substantial part of the events giving rise to the claims occurred in this District. A

19 substantial portion of the affected interstate trade and commerce was carried out in this District.

20 ~~Defendant~~Google has several business units and divisions headquartered in this District, including but

21 not limited to Google AI and Google Research, that are dedicated to either the research, development, or

22 training of ~~Defendants'~~Google's Generative AI Models. ~~Defendants~~Google also ~~employ~~employs

23 hundreds of individuals in this District who are involved in the research, development, and/or training of

24 its Generative AI Models. ~~Defendants have~~Google has transacted business, maintained substantial

25 contacts, and/or committed overt acts in furtherance of the illegal scheme and conspiracy throughout the

26 United States, including in this District. ~~Defendants'~~Google's conduct has had the intended and

27

28

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

1 foreseeable effect of causing injury to persons residing in, located in, or doing business throughout the

2 United States, including in this District.

3   16.   Under Civil Local Rule 3-2(c), assignment of this case to the San Jose Division is proper

4 because this litigation pertains to intellectual property rights, which is deemed a district-wide case

5 category under General Order No. 44. Therefore, venue is proper in any courthouse in this District.

6 **III.   PARTIES**

7   **A. Plaintiffs**

8   17.   **Plaintiff Steve Almond** resides in Massachusetts and is self-employed. He is a New York

9 Times best-selling author whose registered copyrighted works include:

10   • *All the Secrets of the World*;

11   • *Candyfreak: A Journey Through the Chocolate Underbelly of America*;

12   • *The Evil B.B. Chow and Other Stories*;

13   • *Rock and Roll Will Save Your Life: A Book by and for the Fanatics Among Us*;

14   • *Truth Is the Arrow, Mercy Is the Bow: A DIY Manual for the Construction of*

15   *Stories*; and

16   • *Which Brings Me to You: A Novel in Confessions*.

17   18.   Plaintiff Almond is the owner or beneficial owner of registered copyrights in nine (9)

18 literary works, at least six (6) of which—*All the Secrets of the World*, *Candyfreak: A Journey Through*

19 *the Chocolate Underbelly of America*, *The Evil B.B. Chow and Other Stories*, *Rock and Roll Will Save*

20 *Your Life: A Book by and for the Fanatics Among Us*, *Truth Is the Arrow, Mercy Is the Bow: A DIY*

21 *Manual for the Construction of Stories*, and *Which Brings Me to You: A Novel in Confessions*—Google

22 took, used, copied, and/or reproduced without license or authorization (the "Almond Infringed

23 Works"). The registration information for the Almond Infringed Works is contained in **Exhibit A** to this

24 Complaint.

25   19.   The Almond Infringed Works are found on two or more websites, including:

26   • Z-Library (at https://1lib.sk/);

27   • PDF Drive (at https://www.pdfdrive.com/); and

28

1   • OceanofPDF (at https://oceanofpdf.com/).

2   20.   These websites served as sources for Infiniset and C4, training datasets for

3   Gemini. *See* Section IV.

4   21.   Specifically, *All the Secrets of the World* was found on:

5   • Z-Library (at https://1lib.sk/book/21558874/c28c4b/all-the-secrets-of-the-

6   world.html); and

7   • OceanofPDF (at https://oceanofpdf.com/authors/steve-almond/pdf-epub-all-the-

8   secrets-of-the-world-a-novel-download/).

9   22.   *Candyfreak: A Journey Through the Chocolate Underbelly of America* was found on:

10   • Z-Library (at https://1lib.sk/book/17724374/0351fb/candyfreak-a-journey-

11   through-the-chocolate-underbelly-of-america.html);

12   • PDF Drive (at https://www.pdfdrive.com/candyfreak-a-journey-through-the-

13   chocolate-underbelly-of-america-e177466738.html); and

14   • OceanofPDF (at https://oceanofpdf.com/authors/steve-almond/pdf-epub-

15   candyfreak-a-journey-through-the-chocolate-underbelly-of-america-download/).

16   23.   *The Evil B.B. Chow and Other Stories* was found on:

17   • Z-Library (at https://1lib.sk/book/17724383/c01937/the-evil-bb-chow-and-other-

18   stories.html); and

19   • OceanofPDF (at https://oceanofpdf.com/genres/short-stories/pdf-epub-the-evil-b-

20   b-chow-and-other-stories-download/).

21   24.   *Rock and Roll Will Save Your Life: A Book by and for the Fanatics Among Us* was found

22   on:

23   • Z-Library (at https://1lib.sk/book/2208771/11fb64/rock-and-roll-will-save-your-

24   life-a-book-by-and-for-the-fanatics-among-us.html);

25   • PDF Drive (at https://www.pdfdrive.com/rock-and-roll-will-save-your-life-a-

26   book-by-and-for-the-fanatics-among-us-e164900492.html); and

27

28

- OceanofPDF (at https://oceanofpdf.com/authors/steve-almond/pdf-epub-rock-and-roll-will-save-your-life-a-book-by-and-for-the-fanatics-among-us-download-85506150816/).

25.   *Truth Is the Arrow, Mercy Is the Bow: A DIY Manual for the Construction of Stories* was found on:

- Z-Library (at https://1lib.sk/book/28472304/98b8fd/truth-is-the-arrow-mercy-is-the-bow-a-diy-manual-for-the-construction-of-stories.html); and

- OceanofPDF (at https://oceanofpdf.com/authors/steve-almond/pdf-epub-truth-is-the-arrow-mercy-is-the-bow-a-diy-manual-for-the-construction-of-stories-download/).

26.   *Which Brings Me to You: A Novel in Confessions* was found on:

- Z-Library (at https://1lib.sk/book/27459595/49f29b/which-brings-me-to-you.html); and

- OceanofPDF (https://oceanofpdf.com/authors/steve-almond/pdf-epub-which-brings-me-to-you-download/).

27.   The presence of the Almond Infringed Works on the above websites means Google copied and reproduced them multiple times to train Gemini. *See* Section IV.

28.   Plaintiff Almond did not authorize or license Google to use, copy, reproduce, distribute, or train Gemini with the Almond Infringed Works.

29.   In addition to statutory harms, Plaintiff Almond has suffered actual damages through Google's unauthorized copying, reproduction, and use of his works, including but not limited to:

- **Direct Market Harm**: Google's unauthorized copying of the Almond Infringed Works deprived him of licensing revenues he would have received had Google properly licensed his works for AI training purposes.

- **Market Value Diminution**: Google's mass appropriation of literary works has depressed the overall market for fiction and nonfiction writers, like Plaintiff

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

Almond, causing substantial harm to the commercial value of Plaintiff Almond's portfolio of works.

30.   **Plaintiff Sarah Andersen** is domiciled in Oregon and is a self-employed professional cartoonist and illustrator. She is a cartoonist and illustrator whose work includes *Adulthood is a Myth: A "Sarah's Scribbles" 2021 Wall Calendar*, *Adulthood Is a Myth: A Sarah's Scribbles Collection*, *Big Mushy Happy Lump: A Sarah's Scribbles Collection*, *Herding Cats: A Sarah's Scribbles Collection*, and *Oddball: A Sarah's Scribbles Collection*.

31.   Andersen is the owner or beneficial owner of copyrights in at least five (5) works, which Google took, used, copied and/or reproduced without license or authorization (the "Andersen Infringed Works"). The registration information for the Andersen Infringed Works is contained in Exhibit A to this Complaint. The Andersen Infringed Works are contained within the LAION-400M Dataset ("LAION-400M"). *See* Exhibit B to the Complaint, at 2–7. LAION-400M is contained within the LAION-5B dataset.

32.   The presence of the Andersen Infringed Works in LAION-400M means Google copied and reproduced them multiple times to train Imagen. *See* Section IV.

33.   Plaintiff Andersen did not authorize or license Google to use, copy, reproduce, distribute, or train Imagen with the Andersen Infringed Works.

34.   In addition to statutory harms, Andersen has suffered actual damages through Google's unauthorized copying, reproduction, and use of her works, including but not limited to:

- **Direct Market Harm**: Google's unauthorized copying of the Andersen Infringed Works deprived her of licensing revenues she would have received had Google properly licensed her works for AI training purposes.

- **Market Value Diminution**: Google's mass appropriation of artistic works has depressed the overall market for professional illustration and cartooning, causing substantial harm to the commercial value of Andersen's portfolio.

35.   **Plaintiff Burl Barer** resides in Stevenson Ranch, California and is a self-employed author. He is a New York Times best-selling author whose literary work includes:

- *Mom Said Kill*;

- *Broken Doll*;

- *Murder in the Family*;

- *Body Count*;

- *Fatal Beauty*; and

- *Head Shot*.

36.   Plaintiff Barer is the owner or beneficial owner of registered copyrights in six (6) literary works, all of which—*Broken Doll*, *Fatal Beauty*, *Mom Said Kill*, *Headshot*, *Body Count*, and *Murder in the Family*—Google took, used, copied and/or reproduced without license or authorization (the "Barer Infringed Works"). The registration information for the Barer Infringed Works is contained in **Exhibit A** to this Complaint.

37.   The Barer Infringed Works are found on two or more websites, including:

- Z-Library (at https://1lib.sk/);

- Pdfdrive (at https://www.pdfdrive.com/); and

- OceanofPDF (at https://oceanofpdf.com/)).

38.   These websites served as sources for Infiniset and C4, training datasets for Gemini. *See* Section IV.

39.   Specifically, *Broken Doll* was found on:

- Z-Library (at https://1lib.sk/book/4018902/2e9267/broken-doll.html); and

- PDFDrive (at https://www.pdfdrive.com/broken-doll-e196386331.html)).

40.   *Fatal Beauty* was found on:

- Z-Library (at https://1lib.sk/book/1457315/d807c5/fatal-beauty.html)).

41.   *Mom Said Kill* was found on:

- Z-Library (at https://1lib.sk/book/2540922/37e427/mom-said-kill.html)).

42.   *Headshot* was found on:

- Z-Library (at https://1lib.sk/book/1988028/3a3c57/head-shot.html); and

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

- OceanofPDF (at https://oceanofpdf.com/authors/burl-barer/pdf-epub-head-shot-by-burl-barer-download-20591692044/)).

43.   *Body Count* was found on:

- Z-Library (at https://1lib.sk/book/2541254/b8e6a0/body-count.html); and

- PDFDrive (at https://www.pdfdrive.com/body-count-e199614948.html)).

44.   *Murder in the Family* was found on:

- Z-Library (at https://1lib.sk/book/11682997/ab3273/murder-in-the-family.html); and

- OceanofPDF (at https://oceanofpdf.com/authors/burl-barer/pdf-epub-murder-in-the-family-by-burl-barer-download/)).

45.   The presence of the Barer Infringed Works on the above websites means Google copied and reproduced them multiple times to train Gemini. *See* Section IV.

46.   Plaintiff Barer did not authorize or license Google to use, copy, reproduce, distribute, or train Gemini with the Barer Infringed Works.

47.   In addition to statutory harms, Plaintiff Barer has suffered actual damages through Google's unauthorized copying, reproduction, and use of his works, including but not limited to:

- **Direct Market Harm**: Google's unauthorized copying of the Barer Infringed Works deprived him of licensing revenues he would have received had Google properly licensed his works for AI training purposes.

- **Market Value Diminution**: Google's mass appropriation of literary works has depressed the overall market for fiction and nonfiction writers, like Plaintiff Barer, causing substantial harm to the commercial value of Plaintiff Barer's portfolio of works.

48.   **Plaintiff Jessica Fink** is domiciled in New York and is a self-employed professional cartoonist and illustrator. She is a cartoonist and illustrator whose work includes *Chester 5000 XYV, Book 2: Isabelle & George HC*.

49.   Fink is the owner or beneficial owner of copyrights in at least one (1) work, which Google took, used, copied and/or reproduced without license or authorization (the "Fink Infringed Work"). The registration information for the Fink Infringed Work is contained in **Exhibit A** to this Complaint.

50.   The Fink Infringed Work is contained within the LAION-400M Dataset ("LAION-400M"). *See* **Exhibit B** to the Complaint. LAION-400M is contained within the LAION-5B dataset.

51.   The presence of the Fink Infringed Work in LAION-400M means Google copied and reproduced it multiple times to train Imagen. *See* Section IV.

52.   Plaintiff Fink did not authorize or license Google to use, copy, reproduce, distribute, or train Imagen with the Fink Infringed Work.

53.   In addition to statutory harms, Fink has suffered actual damages through Google's unauthorized copying, reproduction, and use of her work, including but not limited to:

- **Direct Market Harm**: Google's unauthorized copying of the Fink Infringed Work deprived her of licensing revenues she would have received had Google properly licensed her work for AI training purposes.

- **Market Value Diminution**: Google's mass appropriation of artistic works has depressed the overall market for professional illustration and cartooning, causing substantial harm to the commercial value of Fink's portfolio.

54.   **Plaintiff Kirsten Hubbard** resides in Los Angeles, California and is a self-employed writer. She is an illustrator and author whose registered copyrighted works include, but are not limited to:

- *Like Mandarin*;

- *Wanderlove*;

- *Watch the Sky*;

- *Race the Night*; and

- *Secrets of Topsea: A Friendly Town That's Almost Always By The Ocean*

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

55.   Plaintiff Hubbard is the owner or beneficial owner of registered copyrights in five (5) written works and one (1) visual art work, all of which—*Like Mandarin*, *Wanderlove*, *Watch the Sky*, *Race the Night*, and *Secrets of Topsea: A Friendly Town That's Almost Always By The Ocean*—Google took, used, copied and/or reproduced without license or authorization (the "Hubbard Infringed Works"). The registration information for the Hubbard Infringed Works is contained in **Exhibit A** to this Complaint.

56.   The Hubbard Infringed Works are found on two or more websites, including:

- Z-Library (at https://1lib.sk/);

- Pdfdrive (at https://www.pdfdrive.com/);

- OceanofPDF (at https://oceanofpdf.com/); and

- Haveibeentrained (at https://haveibeentrained.com/).

57.   Z-Library, Pdfdrive, and OceanofPDF served as sources for Infiniset and C4, training datasets for Gemini; Haveibeentrained is a searchable database of LAION-5B, a training dataset for Imagen. *See* Section IV.

58.   Specifically, *Like Mandarin* was found on:

- Z-Library (at https://1lib.sk/book/1109533/cb641e/like-mandarin.html); and

- OceanofPDF (at https://oceanofpdf.com/authors/kirsten-hubbard/pdf-epub-like-mandarin-download/)).

59.   *Wanderlove* was found on:

- Z-Library (at https://1lib.sk/book/4237162/1cc1e9/wanderlove.html);

- OceanofPDF (at https://oceanofpdf.com/authors/kirsten-hubbard/pdf-epub-wanderlove-download/);

- PDFDrive (at https://www.pdfdrive.com/wanderlove-kirsten-hubbard-e52129165.html); and

  Haveibeentrained (at

  https://haveibeentrained.com/search/TEXT?search_text=wanderlove).

60.   *Wanderlove* also appears within the LAION-400M Dataset ("LAION-400M"). *See* **Exhibit B** to the Complaint.

61.   *Watch the Sky* was found on:

- Z-Library (at https://1lib.sk/book/4827027/e4e67a/watch-the-sky.html); and

- OceanofPDF (at https://oceanofpdf.com/authors/kirsten-hubbard/pdf-epub-watch-the-sky-download-36548863886/)).

62.   *Race the Night* was found on:

- Z-Library (at https://1lib.sk/book/4858313/60315a/race-the-night.html); and

- OceanofPDF (at https://oceanofpdf.com/authors/kirsten-hubbard/pdf-epub-race-the-night-download/)).

63.   *Secrets of Topsea: A Friendly Town That's Almost Always By The Ocean* (published under her pen name Kir Fox) was found on:

- OceanofPDF (at https://oceanofpdf.com/authors/kir-fox/pdf-epub-a-friendly-town-thats-almost-always-by-the-ocean-secrets-of-topsea-1-download/).

64.   The presence of the Hubbard Infringed Works on the above websites and within LAION-400M means Google copied and reproduced them multiple times to train Gemini and Imagen. *See* Section IV.

65.   Plaintiff Hubbard did not authorize or license Google to use, copy, reproduce, distribute, or train Gemini with the Hubbard Infringed Works.

66.   In addition to statutory harms, Plaintiff Hubbard has suffered actual damages through Google's unauthorized copying, reproduction, and use of her works, including but not limited to:

- **Direct Market Harm**: Google's unauthorized copying of the Hubbard Infringed Works deprived her of licensing revenues she would have received had Google properly licensed her works for AI training purposes.

- **Market Value Diminution**: Google's mass appropriation of literary and visual works has depressed the overall market for fiction and nonfiction and visual artists

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

writers, like Plaintiff Hubbard, causing substantial harm to the commercial value
of Plaintiff Hubbard's portfolio of works.

67.   **Plaintiff Hope Larson** is domiciled in North Carolina and is a self-employed illustrator
and author. She has created numerous graphic novels, including *All Summer Long*.

68.   Plaintiff Larson is the owner or beneficial owner of registered copyrights in at least one
(1) literary work that Google took, used, copied and/or reproduced without license or authorization (the
"Larson Infringed Work"). The registration information for the Almond Infringed Works is contained
in **Exhibit A** to this Complaint.

69.   The Larson Infringed Work is found on one or more websites, including Scribd,
https://www.scribd.com.

70.   Scribd was a source for Infiniset and C4, training datasets for Gemini. *See* Section IV.

71.   Specifically, Larson's graphic novel, *All Summer Long*, is found at
https://www.scribd.com/document/753880651/OceanofPDF-com-All-Summer-Long-the-Eagle-Rock-Trilogy-Hope-Larson.

72.   The presence of the Larson Infringed Work on Scribd means Google copied and
reproduced it multiple times to train Gemini. *See* Section IV.

73.   Plaintiff Larson did not authorize or license Google to use, copy, reproduce, distribute, or
train Gemini with the Larson Infringed Work.

74.   In addition to statutory harms, Larson has suffered actual damages through Google's
unauthorized copying, reproduction, and use of her work, including but not limited to:

- **Direct Market Harm**: Google's unauthorized copying and use of Larson's
  graphic novel deprived her of substantial licensing revenues she would have
  received had Google properly licensed her works for AI training purposes.

- **Market Value Diminution**: Google's mass appropriation of literary works has
  depressed the overall market for graphic novel licensing, causing substantial harm
  to the commercial value of Larson's portfolio. This harm is especially acute given

1   the specialized nature of graphic novel creation, which requires extensive creative

2   investment in both visual artistry and narrative development.

3   75.   **Plaintiff Mike Lemos** resides in Moorpark, California and is a self-employed visual

4   artist. He is an illustrator and character designer whose works include, but are not limited to, *Shark Fink*

5   *Sticker*, *Caffeine Case from Outer Space*, and *Butterflies & Snakes.*

6   76.   Lemos is the owner or beneficial owner of copyrights in eight (8) visual art works, at

7   least three of which—*Shark Fink Sticker*, *Caffeine Case from Outer Space*, and *Butterflies & Snakes*—

8   Google took, used, copied and/or reproduced without license or authorization (the "Lemos Infringed

9   Works"). The registration information for the Lemos Infringed Works is contained in **Exhibit A** to this

10   Complaint.

11   77.   The Lemos Infringed Works are contained within the LAION-400M Dataset ("LAION-

12   400M"). *See* **Exhibit B** to the Complaint. LAION-400M is contained within the LAION-5B dataset.

13   78.   The presence of the Lemos Infringed Works in LAION-400M means Google copied and

14   reproduced them multiple times to train Imagen. *See* Section IV.

15   79.   Plaintiff Lemos did not authorize or license Google to use, copy, reproduce, distribute, or

16   train Imagen with the Lemos Infringed Works.

17   80.   In addition to statutory harms, Lemos has suffered actual damages through Google's

18   unauthorized copying, reproduction, and use of his works, including but not limited to:

19   • **Direct Market Harm**: Google's unauthorized copying of the Lemos Infringed

20   Works deprived him of licensing revenues he would have received had Google

21   properly licensed his works for AI training purposes.

22   • **Market Value Diminution**: Google's mass appropriation of illustrated works has

23   depressed the overall market for professional illustrators and character designers,

24   like Plaintiff Lemos, causing substantial harm to the commercial value of his

25   portfolio.

26

27

28

81.   **Plaintiff Jill Leovy** resides in San Antonio, Texas and is a self-employed author. She is a New York Times best-selling author and investigative journalist whose literary work includes *Ghettoside: A True Story of Murder In America.*

82.   Plaintiff Leovy is the owner or beneficial owner of registered copyrights in one (1) literary work—*Ghettoside: A True Story of Murder In America*—which Google took, used, copied and/or reproduced without license or authorization (the "Leovy Infringed Work"). The registration information for the Leovy Infringed Work is contained in **Exhibit A** to this Complaint.

83.   The Leovy Infringed Work is found on two or more websites, including:

- Z-Library (at https://1lib.sk/);

- Pdfdrive (at https://www.pdfdrive.com/);

- OceanofPDF (at https://oceanofpdf.com/); and

- Scribd (at https://www.scribd.com/)).

84.   These websites served as sources for Infiniset and C4, training datasets for Gemini. *See* Section IV.

85.   Specifically, *Ghettoside: A True Story of Murder In America* was found on:

- Z-Library (at https://1lib.sk/book/2753658/688005/ghettoside-a-true-story-of-murder-in-america.html);

- PDFDrive (at https://www.pdfdrive.com/ghettoside-a-true-story-of-murder-in-america-e194586960.html);

- OceanofPDF (at https://oceanofpdf.com/authors/jill-leovy/pdf-epub-ghettoside-a-true-story-of-murder-in-america-download/); and

- Scribd (at https://www.scribd.com/document/254016556/Excerpt-From-Ghettoside-By-Jill-Leovy)).

86.   The presence of the Leovy Infringed Work on the above websites means Google copied and reproduced it multiple times to train Gemini. *See* Section IV.

87.   Plaintiff Leovy did not authorize or license Google to use, copy, reproduce, distribute, or train Gemini with the Leovy Infringed Work.

88.   In addition to statutory harms, Plaintiff Leovy has suffered actual damages through Google's unauthorized copying, reproduction, and use of her work, including but not limited to:

- **Direct Market Harm**: Google's unauthorized copying of the Leovy Infringed Work deprived her of licensing revenues she would have received had Google properly licensed her works for AI training purposes.

- **Market Value Diminution**: Google's mass appropriation of literary works has depressed the overall market for fiction and nonfiction writers, like Plaintiff Leovy, causing substantial harm to the commercial value of Plaintiff Leovy's portfolio of works.

89.   **Plaintiff Connie McLennan** resides in Rocklin, California and is a self-employed artist. She is an author and illustrator whose artwork includes, but is not limited to:

- *The Rainforest Grew All Around*

90.   Plaintiff McLennan is the owner or beneficial owner of registered copyrights in at least one (1) visual artwork——*The Rainforest Grew All Around*——which Google took, used, copied and/or reproduced without license or authorization (the "McLennan Infringed Work"). The registration information for the McLennan Infringed Work is contained in **Exhibit A** to this Complaint.

91.   The McLennan Infringed Work is found on two or more websites, including:

- Z-Library (at https://1lib.sk/);

- Scribd (at https://www.scribd.com/); and

- Haveibeentrained (at https://haveibeentrained.com/).

92.   Z-Library and Scribd served as sources for Infiniset and C4, training datasets for Gemini; Haveibeentrained is a searchable database of LAION-5B, a training dataset for Imagen. *See* Section IV.

93.   Specifically, *The Rainforest Grew All Around* was found on:

- Z-Library (at https://1lib.sk/book/24334626/72ca99/the-rainforest-grew-all-around.html);

- Scribd (at https://www.scribd.com/document/308982138/Rainforest-Preview); and

- Haveibeentrained (at

    https://haveibeentrained.com/search/TEXT?search_text=rainforest%20grew%20all%20around).

94.   *The Rainforest Grew All Around* also appears within the LAION-400M Dataset ("LAION-400M"). *See* **Exhibit B** to the Complaint.

95.   The presence of the McLennan Infringed Work on the above websites and within LAION-400M means Google copied and reproduced it multiple times to train Gemini and Imagen. *See* Section IV.

96.   Plaintiff McLennan did not authorize or license Google to use, copy, reproduce, distribute, or train Gemini with the McLennan Infringed Work.

97.   In addition to statutory harms, Plaintiff McLennan has suffered actual damages through Google's unauthorized copying, reproduction, and use of her works, including but not limited to:

- **Direct Market Harm**: Google's unauthorized copying of the McLennan Infringed Work deprived her of licensing revenues she would have received had Google properly licensed her work for AI training purposes.

- **Market Value Diminution**: Google's mass appropriation of visual works has depressed the overall market for visual artists and illustrators, like Plaintiff McLennan, causing substantial harm to the commercial value of Plaintiff McLennan's portfolio of works.

98.   **Plaintiff Jingna Zhang** is domiciled in Washington state and is a self-employed professional photographer. She is a photographer whose work includes *Anouk* and *Motherland Chronicles*.

99.   Zhang is the owner or beneficial owner of copyrights in at least two (2) photographic works, which Google took, used, copied and/or reproduced without license or authorization (the "Zhang Infringed Works"). The registration information for the Zhang Infringed Works is contained in **Exhibit A** to this Complaint.

100.   The Zhang Infringed Works are contained within the LAION-400M Dataset ("LAION-

1   400M"). *See* **Exhibit B** to the Complaint. LAION-400M is contained within the LAION-5B dataset.

2   101.   The presence of the Zhang Infringed Works in LAION-400M means Google copied and

3   reproduced them multiple times to train Imagen. *See* Section IV.

4   102.   Plaintiff Zhang did not authorize or license Google to use, copy, reproduce, distribute, or

5   train Imagen with the Zhang Infringed Works.

6   103.   In addition to statutory harms, Zhang has suffered actual damages through Google's

7   unauthorized copying, reproduction, and use of her works, including but not limited to:

8   •   **Direct Market Harm**: Google's unauthorized copying of the Zhang Infringed

9   Works deprived her of licensing revenues she would have received had Google

10   properly licensed her works for AI training purposes.

11   •   **Market Value Diminution**: Google's mass appropriation of photographic works

12   has depressed the overall market for professional photography licensing, causing

13   substantial harm to the commercial value of Zhang's portfolio.

14   B.   **Defendant**Defendants

15   104.   **Defendant Alphabet Inc.** Alphabet Inc. ("Alphabet") is a Delaware corporation with its

16   principal place of business at 1600 Amphitheatre Parkway, Mountain View, CA 94043. Alphabet serves

17   as the parent company of Google LLC and exercises direct oversight and control over Google's AI

18   business and development activities. Alphabet derives direct financial benefit from Google LLC's

19   infringing activity.

20   **B.**

21   105.104.   **Defendant Google LLC**. Google LLC is a Delaware corporation with its

22   principal place of business also at 1600 Amphitheatre Parkway, Mountain View, CA 94043. As a

23   subsidiary of Alphabet Inc., Google operates AI-related businesses, including the development and

24   deployment of Generative AI Models and AI-Powered Products. The Generative AI Models at issue in

25   this litigation include all versions, iterations, and relatives of Bard, Gemini, Imagen, PaLM, GLaM,

26   LaMDA, Codey, Chirp, Veo, MedLM, LearnLM, SecLM, Gemma, CodeGemma, RecurrentGemma,

27   PaliGemmiaand LaMDA. The AI-Powered Products include all versions and iterations of Google

28

Search, Google Cloud, Gmail, Google Docs, Google Ads, Google Slides, Chrome, YouTube, Google

Photos, Google Sheets, Google Meet, Google Pixel, Google Maps, Google AI Studio, Google Vids,

Google Workspace, and Vertex AI.

106.105.   **Control and oversight over Google LLC by Alphabet**. Alphabet exercises

substantial control over Google LLC's operations, including its AI development strategy. This control is

evident in Alphabet's direction of AI integration into core Google LLC products such as Google Search,

YouTube, and cloud services. For example, Google Search incorporates Alphabet-owned AI models like

BERT, MUM, and Pathways, reflecting Alphabet's strategic oversight and influence.

107.106.   Additionally, Sundar Pichai serves as both CEO of Alphabet and CEO of Google

LLC, exemplifying integrated leadership across the conglomerate. Under Pichai's leadership, Alphabet

orchestrated the merger of Google Brain with Google DeepMind in 2023 and facilitated the subsequent

transfer of the Gemini team to DeepMind in 2024. These deliberate structural changes illustrate

Alphabet's active role in shaping its subsidiaries' operations to accelerate AI research and development.

Alphabet's chief officers and directors confirm this in their 2023 10-K, stating that Alphabet is

"expanding [its] investment in AI across the entire company. This includes generative AI and continuing

to integrate AI capabilities into our products and services."[13] Alphabet's 2023 10-K goes on to clarify

that "DeepMind is reported as part of Alphabet-level activities[,]" reflecting its "increasing collaboration

with Google Services[. . . .]"[14]

108.107.   Alphabet's overarching control of and direct financial benefit in Google's AI

initiatives is further demonstrated through its substantial investments in entities like DeepMind and other

ventures. This focus on AI as a core strategic priority that extends beyond Google LLC's traditional

business lines—such as search, maps, and YouTube—highlights Alphabet's centralized governance of AI

development across the organization. Alphabet's CEO Sundar Pichai's emphasis on organic AI

development within Alphabet underscores this centralized control, as does his commitment to maintaining

---

[13] *Alphabet Form 10-K*, U.S. SECURITIES AND EXCHANGE COMMISSION (Jan. 30, 2024)
https://www.sec.gov/Archives/edgar/data/1652044/000165204424000022/goog-20231231.htm#i4b6997819c884b839f301ffc0d7fb828_307

[14] *Id.*

the company's autonomy through in-house AI initiatives.[15] . In April of 2024, Pichai reiterated as stated in a document titled "Alphabet Announces First Quarter 2024 Results" announcing Alphabet's revenues and income that "**we are well underway with our Gemini era and there's great momentum across the company. Our leadership in AI research and infrastructure, and our global product footprint, position us well for the next wave of AI innovation**." These actions collectively reflect Alphabet's governance over Google's AI operations and strategic direction. Alphabet has also touted the integration of Gemini across it and Google's suite of products as a major driver of commercial success.

109.108.   The full scope of Alphabet's supervisory authority over and direct financial benefit from Google LLC's infringing conduct remains under investigation, with discovery yet to begin. Upon information and belief, Plaintiffs anticipate that document production, deposition testimony, and expert analysis will reveal substantial additional evidence establishing the prerequisites for vicarious copyright infringement liability.

110.109.   **Unified operations**. Alphabet Inc. and Google LLC function as a seamless, integrated enterprise. Their operations are characterized by intertwined executive leadership and shared financial incentives. Both entities exploit these synergies to profit from the unauthorized use of copyrighted works in building, training, and deploying their AI models, deriving significant commercial gains from these infringing activities.

111.110.   **Agents and co-conspirators.** The unlawful acts alleged against the DefendantsGoogle in this Complaint were authorized, ordered, or performed by the Defendants'Google's respective officers, agents, employees, representatives, or shareholders while actively engaged in the management, direction, or control of the Defendants' businessesGoogle's business or affairs. The

---

[15] *Alphabet Touts Organic AI Development in Q2 Earnings Beat*, PYMNTS (July 23, 2024), https://www.pymnts.com/earnings/2024/alphabet-touts-organic-ai-development-in-q2-earnings-beat/; *see also* Laura Bratton, *Google stock surges on soaring profits and first-ever cash dividend*, QUARTZ (April 25, 2024), https://qz.com/alphabet-google-stock-earnings-gemini-1851436696#:~:text=Google%20parent%20Alphabet%20reported%20a%20surge%20in%20profits,Sundar%20Pichai%20said%20in%20a%20company%20statement%20Thursday. (Alphabet CEO Pichai explaining "**we are well underway with our Gemini era and there's great momentum across the company. Our leadership in AI research and infrastructure, and our global product footprint, position us well for the next wave of AI innovation**.").

1   Defendants'Google's agents operated under the explicit and apparent authority of their principals. Each

2   GoogleDefendant, and its subsidiaries, affiliates, parents, and agents operated as a single unified entity.

3   112.111.   Various persons or firms not named as Defendantsa Defendant may have

4   participated as co-conspirators in the violations alleged herein and may have performed acts and made

5   statements in furtherance thereof. Each acted as the principal, agent, or joint venture of, or for other

6   Defendantsa Defendant with respect to the acts, violations, and common course of conduct alleged herein.

7   **IV.   FACTUAL ALLEGATIONS**

8   **A.   The Generative AI arms race**

9   113.112.   The term "generative artificial intelligence" or "generative AI" refers to systems

10   or models capable of producing output that simulates human expression, often in response to user inputs

11   called "prompts." In the race to dominate this emerging field, companies like Google have "shifted their

12   entire corporate strategies in order to seize control of what they believe will become a new infrastructure

13   layer of the economy."[16] Staying ahead of the ever-growing pack has become paramount. Google itself

14   has declared a "code red," acknowledging that "the company may be approaching a moment that the

15   biggest Silicon Valley outfits dread — the arrival of an enormous technological change that could upend

16   the business."[17]

17   114.113.   In response, Google has directed unprecedented resources toward generative AI.

18   Google has consolidated its AI research divisions, such as DeepMind and Google Brain, on the belief

19   that generative AI will define the next era of technology.[18] Google has already integrated generative AI

20   throughout its products, betting on it as the foundation for future growth. This effort represents not just a

21   technological shift but a strategic overhaul, as Google seeks to dominate a field poised to reshape

22   industries and redefine how humans interact with technology.

---

[16] Andrew R. Chow, *The AI Arms Race is Changing Everything*, TIME (Feb. 16, 2023), https://time.com/magazine/us/6256547/february-27th-2023-vol-201-no-7-u-s/.

[17] Nico Grant & Cade Metz, *A New Chat Bot is a "Code Red" for Google's Search Business*, N.Y. TIMES (Dec. 21, 2022), https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html.

[18] *Google Consolidates Its DeepMind and Research Teams Amid AI Push*, REUTERS (Apr. 18, 2024), https://www.reuters.com/technology/google-consolidates-its-deepmind-research-teams-amid-ai-push-2024-04-18/.

i. Generative AI Model types

115.114.   Relevant to this action are three types of generative AI models: large language models ("LLMs"), text-to-image diffusion models ("T2I models"), and multimodal large language models ("MLLMs").

116.115.   LLMs, such as the early versions of Gemini and Bard, accept user text prompts as input text and generate responsive text as output. An LLM's ability to produce coherent and convincing responses to prompts depends heavily on the size and quality of the datasets used for its training.

117.116.   Training an LLM first requires copying a massive set of data files, called a *training dataset*. The training dataset then undergoes multiple preparatory stages processing, most of which cause more copies to be made. Finally, copies of the training dataset are ingested by the model during training. After the initial round of training, a model may be fine-tuned to improve performance on specific tasks—a separate training process that often involves other training datasets, copied from other sources—and finally prepared for deployment.

118.117.   T2I models, like Imagen, use text prompts to create image outputs through a machine-learning technique called diffusion. T2I models are trained by copying an enormous quantity of digital images with associated text captions, extracting protected expression from these works, and transforming the protected expression into a large set of numbers called weights that are stored within the model. The weights are entirely and uniquely derived from the protected expression in the training dataset. When a diffusion model generates an image in response to a user prompt, it performs a computation that relies on these stored weights with the purpose of imitating the protected expression ingested from the training dataset.

119.118.   MLLMs, like the current version of Gemini, are LLMs that accept input and generate output in multiple formats. MLLMs are built with training datasets that comprise of a variety of text, image, audio, and video data. Like LLMs and T2I models, the training of MLLMs requires making multiple copies of the works in the training dataset—including the protected expression within those works.

120.119.   Google made unauthorized copies of copyrighted works—including the Plaintiff Works—to build Generative AI Models that generate content depressing the overall market for those works. As part of this process, Google copied the Plaintiff Works multiple times: first a copy is made during data collection, then a series of copies are made during model training, and finally a copy is permanently incorporated into the parameters of the Generative AI Model. Google monetizes its copyright infringements through two revenue streams: (1) direct revenue from subscription and licensing fees for access to its Generative AI Models, and (2) indirect revenue from integration of those Generative AI Models into its AI-Powered Products. This comprehensive exploitation of the Plaintiff Works demonstrates both the commercial nature of Google's infringement and the immediacy of harm to Plaintiffs' rights.

ii.   Google willfully trained Gemini on copyrighted works

121.120.   Gemini's predecessor, Bard, was launched in March 2023. Google told consumers that Bard was able to perform a variety of content creation tasks, such as writing stories and drafting personal or commercial content, including in multiple tones and styles.[19]

122.121.   Google developed Bard using an LLM called LaMDA.[20]

123.122.   LaMDA was developed and trained on Google's Infiniset Dataset ("Infiniset"), a vast 1.56 trillion-word corpus of internet content that incorporates the C4 Dataset ("C4").[21] C4, created by Google in 2020, is a filtered version of the Common Crawl dataset.[22] Infiniset is comprised of 50% of dialogues from public forums (e.g. Reddit, Twitter), 12.5% C4 dataset, 12.5% code documents (e.g. GitHub, Stack Overflow)), 12.5% Wikipedia, 6.5% English web documents, and 6.5% non-English web documents.[23]

---

[19] https://blog.google/products/gemini/how-to-use-google-bard/ (last accessed December 16, 2024).

[20] Aditya Sharma, *Image Processing with Gemini Pro*, PY IMAGE SEARCH (Feb. 12, 2024), https://pyimagesearch.com/2024/02/12/image-processing-with-gemini-pro/.

[21] Roger Montti, *Google Bard AI—What Sites Were Used To Train It*, SEARCH ENGINE J. 4 (Feb. 10, 2023), https://www.searchenginejournal.com/google-bard-training-data/478941/.

[22] *Dataset Card for C4*, HUGGING FACE, https://huggingface.co/datasets/legacy-datasets/c4 (last visited Nov. 22, 2024).

[23] Sakil Ansari, *The Rise of Google Bard*, Medium (Mar. 25, 2023), https://sakilansari4.medium.com/the-rise-of-google-bard-4d43225c18e4

1   ~~124.~~123.   Common Crawl is the largest freely available collection of web crawl data and

2   one of the most important sources of pre-training data for large language models."[24] Its web

3   archive consists of petabytes of data collected since 2008.[25] It completes crawls of the internet and

4   releases its web archives generally every month.[26]

5   ~~125.~~124.   C4 is a vector for copyright infringement. The copyright symbol appears many

6   times within C4, indicating the presence of numerous copyrighted works within the dataset. Upon

7   information and belief, Google was never authorized to make copies of any copyrighted works within

8   C4.[27]

9   ~~126.~~125.   C4 contains massive quantities of copyrighted materials, including works found

10   on piracy sites or nonconsenting digital libraries, and through subscription services. For example, C4

11   contains data from "b-ok.org," a "notorious market for pirated e-books," as well as "[a]t least 27 other

12   sites identified by the U.S. government as markets for piracy and counterfeits."[28] The website "b-

13   ok.org," also known as "Z-Library," is "[t]he world's largest [illegal] ebook library and digital library."[29]

14   In 2022, the site's operators faced criminal prosecution for "criminal copyright infringement, wire fraud,

---

[24] Stefan Baack, *A Critical Analysis of the Largest Source for Generative AI Training: Common Crawl* at 1, MOZILLA FOUND. (June 3, 2024), https://facctconference.org/static/papers24/facct24-148.pdf.

[25] Rosanna Xia, *Tech Entrepreneur Gil Elbaz Made it Big in L.A.*, L.A. TIMES (February 5, 2012), https://www.latimes.com/business/la-xpm-2012-feb-05-la-fi-himi-elbaz-20120205-story.html.

[26] *Statistics of Common Crawl Monthly Archives*, COMMON CRAWL, https://commoncrawl.github.io/cc-crawl-statistics/ (last visited Dec. 10, 2024).

[27] Kevin Schaul, *et. al.*, *Inside the Secret List of Websites that Make AI Like ChatGPT Sound Smart*, WASH. POST (April 19, 2023), https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/.

[28] *See id.*

[29] Beinginstructor, ZLibrary — *The World's Largest Ebook Library*, MEDIUM (Feb. 16, 2023), https://medium.com/@beinginstructor/zlibrary-the-worlds-largest-ebook-library-dfb933762cfc.

and money laundering."[30] As another example, the pirated book site "oceanofpdf.org" also appears in C4.[31, 32]

~~127.~~126.　　　The third largest site within C4 is Scribd.com ("Scribd"), a subscription-based digital library with 60 million e-books and audio books that compensates authors "for their content through a revenue-sharing model based on the number of reads their work receives from subscribers."[33]

~~128.~~127.　　　Google's Infiniset (which encompassed C4) was a foundational dataset used to train Bard, Gemini's predecessor. Upon information and belief, Google used Infiniset and C4 to train Gemini.

~~129.~~128.　　　Unauthorized copies of the copyrighted works of Plaintiffs Steve Almond, Burl Barer, Kirsten Hubbard, Jill Leovy, and Connie McLennan are contained within Infiniset and C4. On information and belief, unauthorized copies of the Plaintiff Hope Larson's copyrighted works are contained within Infiniset and C4. Google's use of Infiniset and C4 to build and train Gemini constitutes direct copyright infringement.

~~130.~~129.　　　Google has not revealed the sources of the rest of its training corpora. Google for example has only identified other training data as "web documents." On information and belief, Google likely used other datasets containing copyrighted works to train Gemini.

~~131.~~130.　　　Plaintiffs' investigation of Google's willful infringement remains ongoing~~, with discovery recently yet to begin.~~. Upon information and belief, Plaintiffs anticipate that discovery will reveal substantial additional evidence of willfulness, in addition to further evidence of the presence of

---

[30] *Pirate Website Z-Library Taken Down and Alleged Operators Arrested*, INT'L PUBLISHERS ASS'N (Nov. 19, 2022), https://internationalpublishers.org/pirate-website-z-library-taken-down-and-alleged-operators-arrested/.

[31] OceanofPDF has operated under various domain names, including oceanofpdf.com and oceanofpdf.org. These domain names have been associated with the same platform, which offer pirated books.

[32] *Actions Authors Can Take Against Notorious Piracy Site OceanofPDF.com*, AUTHORS GUILD (Nov. 22, 2024), https://authorsguild.org/news/oceanofpdf-piracy-site-actions-authors-can-take/.

[33] Mihaela Bidilică, *How to Publish on Scribd: A Simple Guide to Scribd Publishing*, PUBLISHDRIVE (Apr. 26, 2023), https://publishdrive.com/publishing-on-scribd-the-complete-review-for-authors.html.

Plaintiffs' Works in Gemini training datasets because per Google's own research, books have been a vital and large part of its LLM training datasets.[34]

iii.   Google willfully trained Imagen on copyrighted works

~~132.~~131.   In May 2022, Google announced Imagen in a paper called "Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding."[35] In the paper, Google admits that it trained Imagen on "the publicly available Laion [sic] dataset . . . with ≈ 400M image-text pairs."[36] This dataset is referred to as LAION-400M and was also used to train Imagen 2 and Imagen 3.

~~133.~~132.   In November 2022, Google made Imagen publicly available to a select group of users through its AI Test Kitchen app. According to reporting at the time, Google "announced it will be adding Imagen—in a *very* limited form—to its AI Test Kitchen app as a way to collect early feedback on the technology."[37]

~~134.~~133.   In May 2023, Google made Imagen even more widely available through its commercial AI cloud-computing service, called Vertex AI. According to a Google blog post about Vertex AI, Google described it as "Imagen, our text-to-image foundation model, lets organizations generate and customize studio-grade images at scale for any business need."[38]

~~135.~~134.   In October 2023, Google made Imagen even more widely available through a tool called Search Generative Experience. According to reporting at the time, "If you're opted in to [Search Generative Experience] through Google's Search Labs program, you can just type your query into the

---

[34] *See, e.g., PaLM: Scaling Language Modeling with Pathways* (October 5, 2023), https://arxiv.org/pdf/2204.02311, and *GLaM: Efficient Scaling of Language Models with Mixture-of-Expert* (August 1, 2022) https://arxiv.org/pdf/2112.06905.

[35] Chitwan Saharia, *Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding* at 7, GOOGLE RES. (May 23, 2022), https://arxiv.org/abs/2205.11487.

[36] *Id.*

[37] *See* https://www.theverge.com/2022/11/2/23434361/google-text-to-image-ai-model-imagen-test-kitchen-app

[38] *See* https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-launches-new-ai-models-opens-generative-ai-studio

1   Google search bar. After you do, [Search Generative Experience] can create a few images based on your

2   prompt that you can pick from. The tool is powered by the Imagen family of AI models."[39]

3   ~~136.~~135.   In December 2023, Google released the successor to Imagen, called Imagen 2.

4   Unlike the paper that accompanied the initial version of Imagen, Google's introduction of Imagen 2

5   carefully omits a detailed description of its training dataset. Google limits itself to vague comments such

6   as "From the outset, we invested in training data safety for Imagen 2, and added technical guardrails to

7   limit problematic outputs like violent, offensive, or sexually explicit content."[40]

8   ~~137.~~136.   The creation and deployment of the Imagen models underscore Google's ambition

9   to dominate the generative AI market by providing tools that reshape how creative content is produced.

10 However, this success has come at the expense of copyright holders whose works were copied without

11 authorization to train Imagen. By incorporating vast numbers of copyrighted works within its training

12 datasets, Google built Imagen on a foundation of misappropriated intellectual property—namely, the

13 contents of LAION-400M dataset—directly profiting from the unauthorized copying of others' creative

14 efforts.

15   **a.   LAION-400M contains Plaintiffs' copyrighted works**

16   ~~138.~~137.   LAION (acronym for "Large-Scale Artificial Intelligence Open Network") is an

17 organization based in Hamburg, Germany. LAION's most well-known projects are the datasets of

18 training images, like LAION-400M, that it has released publicly for training models. These datasets are

19 widely used in the AI industry and free for use.

20   ~~139.~~138.   LAION publicly released LAION-400M, a dataset of 400 million training images

21 assembled from images accessible on the internet, in August 2021. Romain Beaumont, a chief architect

22 at LAION, also created and released an online tool called "Clip Retrieval" that acted as a search

23 interface to LAION to check whether certain artists or artworks were included in the LAION-400M

24 dataset.[41]

25

26 [39] *See* https://www.theverge.com/2023/10/12/23913337/google-ai-powered-search-sge-images-written-drafts

27 [40] *See* https://deepmind.google/technologies/imagen-2/

28 [41] *See Clip Retrieval*, https://rom1504.github.io/clip-retrieval (last visited Nov. 22, 2024).

1   140.139.   A download of the LAION-400M dataset also includes a list of metadata records,

2   one for each training image. Each record includes the URL of the image, the image caption, a

3   measurement of the similarity of the caption and image, a NSFW flag (indicating the probability the

4   image contains so-called "not safe for work" content), and the width and height of the image.

5   141.140.   The actual images referenced in the LAION-400M dataset records are not

6   included with the dataset. Instead, anyone, such as Google, who wished to use LAION-400M for

7   training their own model needed to first acquire copies of the actual images from the URLs listed in the

8   metadata records. To facilitate the copying of these images, Beaumont created a software tool called

9   "img2dataset" that takes the LAION-400M metadata records as input and makes copies of the

10   referenced images from the URLs in each metadata record, thereby creating local copies for the user.

11   The "img2dataset" tool is distributed from a page controlled by Beaumont on GitHub.[42] LAION also

12   promotes the `"img2dataset`" tool in its documentation for LAION-400M: "This metadata dataset

13   purpose is to download the images for the whole dataset or a subset of it by supplying it to the very

14   efficient `img2dataset`'img2dataset' tool."[43]

15   142.141.   Training a model on the LAION-400M dataset is infeasible without first using

16   `"img2dataset`" or another similar tool to download the images from the URLs in the dataset. In other

17   words, when Google trained Imagen on LAION-400M, as it did, it first and necessarily had to make one

18   or more copies of each image. Then, when training the model, Google made more unauthorized copies

19   and reproduction of the images.

20   143.142.   Plaintiffs Jingna Zhang, Sarah Andersen, Jessica Fink, Mike Lemos, and Connie

21   McLennan, each own one or more copyrighted works found within LAION-400M. They never

22   authorized Google to copy their images for training any of Google's Generative AI Models, including

23   Imagen.

24   **b.   LAION-5B and Imagen 2 and 3**

25

26   _____

27   [42] *rom 1504 / img2dataset*, GITHUB, https://github.com/rom1504/img2dataset (last visited Nov. 22, 2024).
     [43] *See* Christoph Schuhmann, *LAION-400-Million Open Dataset*, LAION (Aug. 20, 2021),

28   https://laion.ai/blog/laion-400-open-dataset/.

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

144.143.   On March 31, 2022, LAION released LAION-5B, a dataset of 5.85 billion training images—14 times more images than contained in LAION-400M. Beaumont, who was hired as an engineer at Google, authored the LAION blog post announcing LAION-5B's launch.[44] On information and belief, Google hired Beaumont primarily to influence the creation of future LAION image datasets based on a) Beaumont's key role creating LAION-400M—which Google used to train Imagen; b) Beaumont's control of the "img2dataset" tool that was essential to using the LAION-400M dataset, and c) Beaumont's control of the "Clip Retrieval" website that was essential to searching the LAION-400M dataset.

145.144.   In August 2022, Beaumont created a specialized AI model to rate the aesthetic quality of an image and used this model to create subsets of the LAION-5B training images filtered by aesthetic quality. Beaumont called this model "LAION-Aesthetics." In its introduction of Imagen 2 in December 2023, Google said, "[w]e trained a specialized image aesthetics model based on human preferences for qualities like good lighting, framing, exposure, sharpness, and more. Each image was given an aesthetics score which helped condition Imagen 2 to give more weight to images in its training dataset that align with qualities humans prefer."[45] On information and belief, Beaumont's work on LAION-Aesthetics formed the basis of Imagen 2's "aesthetics model" since at the time Beaumont was both a contributor to LAION and a full-time employee of Google.

146.145.   In October 2022, Romain Beaumont was a primary author of "LAION-5B: An open large-scale dataset for training next generation image-text models" (hereinafter, the "Beaumont–LAION-5B Paper"). According to the Beaumont–LAION-5B Paper, LAION-400M is a subset of LAION-5B, meaning every image in LAION-400M is also in LAION-5B.

147.146.   Like the LAION-400M dataset, the actual images referenced in the LAION-5B dataset records are not included with the dataset. Anyone who wishes to use LAION-5B for training their own model must also first acquire copies of the actual images from their URLs either using "img2dataset" or a similar tool.

---

[44] *See* Romain Beaumont, *LAION-5B: A New Era of Open Large-Scale Multi-Modal Datasets*, LAION (Mar. 31, 2022), https://laion.ai/blog/laion-5b/.

[45] *See* *Imagen 2*, GOOGLE, https://deepmind.google/technologies/imagen-2/ (last visited Nov. 22, 2024).

148.147.   LAION-5B was likely used by Google to train Imagen 2.

149.148.   Imagen 3, released in August 2024, represents the latest iteration of Google's T2I model, showcasing even greater sophistication and precision in generating high-quality visuals. Based on the methods and datasets used to train its predecessors, as detailed above, it is highly likely that Google relied on LAION-5B to train Imagen 3. This reliance would have involved the unauthorized use of copyrighted works contained within LAION-5B, continuing the pattern of infringement that has defined the development of the Imagen series.

150.149.   Plaintiffs Jingna Zhang, Sarah Andersen, Jessica Fink, Connie McLennan, and Mike Lemos own one or more copyrighted works found within LAION-5B. They never licensed or authorized Google to copy their images from LAION-5B or use them for training any model.

iv.   Google's infringement was willful

151.150.   Google's unauthorized copying of Plaintiff Works in the development of its Generative AI Models, including Gemini and Imagen, was willful. Google knowingly copied copyrighted works into its training datasets without securing licenses or authorization, demonstrating a flagrant disregard for U.S. copyright laws.

152.151.   Google was well aware that the training datasets it copied, such as LAION-400M and C4, contained copyrighted works. These datasets included works from known piracy sources, subscription-based platforms, and publicly flagged copyright-protected works. For example, the presence of copyright symbols within the C4 dataset should have served as an unmistakable indicator of the protected status of the included works. Despite this, Google proceeded to use these datasets without authorization from the copyright holders.

153.152.   Google was also well aware that potential and current customers were concerned about using its Generative AI Models, since they were built and trained by infringing upon the exclusive rights of copyright owners. Rather than change its conduct, Google rolled out an indemnification program for its Generative AI Model customers that stated, "[i]f you are challenged on copyright grounds, we will assume responsibility for the potential legal risks involved." Rather than address its

1   infringing conduct directly, Google used its financial strength as a means to actively encourage the use

2   of its Generative AI Models.

3   154.153.   Google's executives and developers were aware of the legal risks associated with

4   using copyrighted works in training their AI models. Public statements by Google acknowledge the

5   importance of respecting copyright, yet Google chose to prioritize speed to market and dominance in the

6   generative AI space over compliance with copyright laws.

7   155.154.   Google's decision to embed its unlawfully trained Generative AI Models into its

8   AI-Powered Products further underscores the willful nature of its infringement. By incorporating

9   Generative AI Models like Gemini and Imagen into AI-Powered Products like Google Cloud, YouTube,

10   and Google Ads, Google not only infringed the copyrighted works but also actively monetized them,

11   generating billions in revenue.

12   156.155.   The scale and systematic nature of Google's infringement foreclose any claim of

13   good faith or ignorance. Google's immense resources, sophisticated technological capabilities, and legal

14   expertise make its failure to comply with copyright laws particularly egregious. Google's continued

15   refusal to acknowledge or remedy its infringement, even as it reaps massive revenue from its AI models,

16   demonstrates willfulness and a conscious decision to prioritize revenue over creators' rights.

17   157.156.   Plaintiffs and other copyright holders have suffered significant and ongoing harm

18   as a direct result of Google's willful infringement. By copying their works without authorization, Google

19   deprived them of licensing revenue, diminished the value of their copyrights, and irreversibly entangled

20   their copyrighted works in the company's AI-Powered Products.

21   158.157.   Defendants haveGoogle has also engaged in a knowing and active fraudulent

22   concealment and denial of the facts alleged through the time period relevant to this action.

23   159.158.   At all times relevant to this Complaint, Defendants wereGoogle was aware of

24   material facts that were necessary to the Plaintiffs' understanding of whether their exclusive rights under

25   the Copyright Act were violated. These facts included, but were not limited to, DefendantsGoogle's

26   concealment of the data and datasets upon which the Generative AI Models were trained, whether the

27   Plaintiffs works were used to train the Generative AI Models, and how the Plaintiffs Works were used,

28

1   copied, and/or reproduced in the training process of the Generative AI Models or otherwise used for

2   integration into the AI-Powered Products.

3   ~~160.~~159.   Plaintiffs did not have access to the inner workings and strategies of Google,

4   including those related to the building and training of its Generative AI Models, the creation of training

5   datasets that contained their works, or how the Generative AI Models were being integrated into the AI-

6   Powered Products.

7   ~~161.~~160.   ~~Defendants~~Google intentionally concealed these material facts with the intent to

8   mislead, deceive, or defraud Plaintiffs into taking actions or refraining from taking actions that Plaintiffs

9   otherwise would not have taken.

10   ~~162.~~161.   Plaintiffs did not know, and could not have reasonably known, about the

11   concealed facts despite exercising reasonable diligence. Plaintiffs relied on ~~the Defendants'~~Google's

12   representations and omissions, which were fraudulent and intended to induce Plaintiffs to act to its

13   detriment.

14   **V.   CLASS ALLEGATIONS**

15   ~~163.~~162.   The "**Class Period**" as defined in this Complaint begins on or at least as early as

16   January 1, 2017 and runs through the present. Because Plaintiffs do not yet know when the unlawful

17   conduct alleged herein began, but believe, on information and belief, that the conduct likely began earlier

18   than the date listed above, Plaintiffs reserve the right to amend the Class Period to comport with the facts

19   and evidence uncovered during further investigation or through discovery.

20   ~~164.~~163.   **Class definition**. Plaintiffs bring this action for damages and injunctive relief as a

21   class action under Federal Rules of Civil Procedure 23(a), 23(b)(2), and 23(b)(3), on behalf of the

22   following Class:

23   All persons or entities ~~who: (1) are~~ domiciled in the United States~~; (2) own~~
who owned a ~~valid~~United States copyright ~~registration for one or more works~~
~~under the Copyright Act; (3) whose exclusive rights under 17 U.S.C. § 106~~ in
24   ~~their registered works were infringed upon, under 17 U.S.C. § 501,~~any work
used by Google ~~without license or authorization in order~~ to train Google's
25   Generative AI Models during the Class Period~~; and (4) held such copyright~~
~~registration prior to Google's unauthorized use~~.
26

27   ~~165.~~164.   This Class definition excludes:

28

a.   ~~Defendants~~The Defendant named herein;

b.   any of ~~the Defendants'~~Google's co-conspirators;

c.   any of ~~Defendants'~~Google's parent companies, subsidiaries, and affiliates;

d.   any of ~~Defendants'~~Google's officers, directors, management, employees, subsidiaries, affiliates, or agents;

e.   all governmental entities; and

f.   the judges and chambers staff in this case, as well as any members of their immediate families.

~~166.~~165.   **Numerosity**. The exact number of members in the Class is in the exclusive control of ~~Defendant~~Google, but on information and belief there are at least thousands of members in the Class geographically dispersed throughout the United States. Therefore, joinder of all members of the Class in the prosecution of this action is impracticable.

~~167.~~166.   **Typicality**. Plaintiffs' claims are typical of the claims of other members of the Class because Plaintiffs and all members of the Class were damaged by the same wrongful conduct of ~~Defendants~~Google as alleged, and the relief sought is common to all members of the Class.

~~168.~~167.   **Adequacy**. Plaintiffs will fairly and adequately represent the interests of the members of the Class because the Plaintiffs have experienced the same harms as the members of the Class and have no conflicts with any other members of the Class. Furthermore, Plaintiffs have retained sophisticated and competent counsel who are experienced in prosecuting federal and state class actions, as well as other complex litigation.

~~169.~~168.   **Commonality and predominance**. Numerous questions of law or fact common to each Class Member arise from ~~Defendants'~~Google's conduct and predominate over any questions affecting the members of the Class individually:

- 169.   Whether ~~Defendants~~Google violated Plaintiffs' and Class Members' exclusive rights under the Copyright Act when they took, copied, and used Plaintiffs' copyrighted works without license or authorization and used them to build and train the Google's Generative AI Models.

- Whether any affirmative defense excuses ~~Defendants'~~Google's conduct.

- Whether any statutes of limitation constrain the potential for recovery for Plaintiffs and the Class;

- Whether ~~Defendants'~~Google's conduct constitutes an infringement of the copyrights owned by Plaintiffs and the Class in their respective works;

- Whether ~~Defendants'~~Google's copying and/or reproduction of the copyrighted works constitutes copyright infringement;

- Whether ~~Defendants'~~Google's copying and/or reproduction of the copyrighted works constitutes fair use;

- Whether ~~Defendants'~~Google's violation of Plaintiffs' and the Class's exclusive rights under copyright law entitles them to damages, including statutory damages, and the amount of statutory damages; and

- Whether ~~Defendants~~Google acted willfully with respect to the copyright infringement.

170.   **Other class considerations**. ~~Defendants'~~Google's actions broadly affect the Class, making this class action the most efficient means of resolving the controversy. Prosecuting the claims pleaded as a class action will eliminate the possibility of repetitive litigation. There will be no material difficulty in the management of this action as a class action. The prosecution of separate actions by individual Class members would create the risk of inconsistent or varying adjudications, establishing incompatible standards of conduct for ~~Defendants~~Google.

VI.   CAUSES OF ACTION

<div align="center">

COUNT 1

Direct Copyright Infringement (17 U.S.C. § 501)

Against Google LLC

</div>

171.   The preceding factual allegations are incorporated by reference.

172.   Plaintiffs and Class Members are the legal owners of valid copyright registrations in their respective works, having satisfied all statutory formalities required by the Copyright Act.

173.   The Copyright Act grants copyright owners exclusive rights enumerated in 17 U.S.C. § 106, including but not limited to: a) the right to reproduce the copyrighted works in copies; b) the right

1   to prepare derivative works; c) the right to distribute copies to the public; and d) the right to display the

2   copyrighted works publicly.

3   174.   Google ~~LLC~~ directly infringed upon Plaintiffs' exclusive rights through multiple distinct

4   acts of unauthorized reproduction and use: a) copying of works when assembling training datasets

5   through tools like ~~'img2dataset';~~"img2dataset"; b) multiple reproductions during the pre-training and

6   training phases of model development; and/or c) creating derivative works when assembling training

7   datasets or during the pre-training and training phases of model development.

8   175.   Each unauthorized copy or use constitutes a separate and distinct act of infringement

9   under 17 U.S.C. § 501. ~~Google LLC's~~Google's systematic copying was deliberate and commercial in

10 nature, and executed to develop and monetize its Generative AI Models.

11   176.   ~~Google LLC's~~Google's infringement has directly resulted in at least two distinct forms of

12 monetary harm: a) lost licensing revenue from Google ~~LLC~~ that Plaintiffs would have received had

13 Google ~~LLC~~ properly obtained authorization; and b) loss of value to Plaintiffs' respective portfolios due

14 to ~~Google LLC's~~Google's mass infringement that has depressed the overall market for the Plaintiffs'

15 respective commercial markets.

16   177.   ~~Defendants~~Google deliberately copied huge quantities of copyrighted works, knowing

17 these works would form the foundation of their commercial AI products. The fact that the training

18 datasets contained copyrighted works is not a secret. For example, it is well-known that the LAION

19 datasets contained copyrighted images. Further, the C4 dataset contains numerous instances of the

20 presence of the copyright symbol which denotes a copyrighted work. ~~Defendants'~~Google's conscious

21 decision to proceed without obtaining licenses, despite clear indications of copyright protection,

22 demonstrates the willful nature of ~~their~~its infringement.

23   178.   ~~Defendants'~~Google's infringing conduct directly advances ~~their~~its commercial interests.

24 By misappropriating Plaintiffs' copyrighted materials, ~~Defendants~~Google avoided substantial licensing

25 costs while accelerating development of their AI products. These products now generate significant

26 revenue through both direct monetization and integration into ~~Google LLC's~~Google's broader product

27 suite and AI-Powered Products.

28

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

179.   ~~Google LLC's~~Google's infringement was willful, as demonstrated by: a) deliberate copying from datasets known to contain copyrighted material; b) use of content from known piracy sites; c) failure to implement any meaningful copyright compliance measures; and d) continued exploitation despite knowledge of ownership claims.

180.   Plaintiffs and Class Members are entitled to: a) statutory damages of up to $150,000 per work for willful infringement or actual damages and any profits of ~~the Defendants~~Google that are attributable to the infringement, pursuant to 17 U.S.C. § 504; b) injunctive relief under 17 U.S.C. § 502; and c) recovery of full costs and reasonable attorney's fees under 17 U.S.C. § 505.

181.   Plaintiffs and Class Members will suffer immediate and irreparable injury absent permanent injunctive relief. The unique characteristics of AI model training and deployment create distinct categories of harm that resist traditional monetary remediation. Further, Google's infringement ~~in~~is continuing, and monetary damages cannot prevent the threat of future infringement.

~~COUNT 2~~

~~Vicarious Copyright Infringement~~

~~against Alphabet~~

~~182.   The preceding factual allegations are incorporated by reference.~~

~~183.   Vicarious liability for copyright infringement requires: a) the right and ability to supervise or control the infringing activity; b) a direct financial interest in such activities, and c) a failure to exercise the right and ability to supervise or control the infringing activity. Alphabet satisfies all three elements, through its relationship with Google and its active role in the infringing conduct.~~

~~184.   Alphabet maintained both the legal right and practical ability to supervise Google's infringing activities through: a) integrated executive leadership, including shared CEO Sundar Pichai who directs both entities' strategic operations; b) direct operational control over Google's AI initiatives, including Alphabet's orchestration of the Google Brain-DeepMind merger fundamental to building and training the Generative AI Models; c) oversight of Google's technological development through shared executive officers, including the Chief Financial Officer and Chief Investment Officer; d) strategic direction of AI integration across Google's product portfolio; and e) authority to implement copyright~~

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

1   compliance measures, which Alphabet chose not to exercise despite knowledge of systematic

2   infringement.

3   185.   Alphabet derived direct financial benefits from Google's infringing conduct through

4   multiple channels: a) increased market capitalization following the release of AI products built on

5   infringing materials; b) revenue growth from Google Cloud's AI-driven services, reaching $10.9 billion

6   quarterly; c) enhanced competitive position in the AI market secured through accelerated development

7   enabled by infringement; d) cost savings from avoiding necessary licensing fees for copyrighted works;

8   and e) integration of infringing AI models across Google's product suite, driving increased usage and

9   revenue.

10   186.   The scale of financial benefit is evidenced by: a) record quarterly revenues of $88.3

11   billion attributed to AI integration; b) 29% year-over-year growth in Google Cloud revenue directly

12   linked to AI capabilities; c) billions in revenue generated from generative AI solutions within the first

13   year; d) market capitalization growth to $2.05 trillion,[46] representing a 16.5% increase following AI

14   product releases in the fourth quarter of 2023;[47] and e) enhanced valuation metrics directly attributable

15   to AI market position.

16   187.   Google has derived substantial profits directly attributable to its unauthorized use of

17   Plaintiffs' copyrighted works. The integration of Google's generative AI models, such as Gemini and

18   Imagen, into its broader product ecosystem — including Google Cloud, Gmail, Google Docs, YouTube,

19   and Google Ads, among other offerings — has significantly enhanced the company's revenue streams. In

20   2024, Google reported record-breaking quarterly revenues of $88.3 billion, with CEO Sundar Pichai

21   crediting the success of AI-driven innovations as a key contributor to this financial achievement.

22   Specifically, Google Cloud, powered by generative AI solutions, generated $10.9 billion in a single

23   quarter, marking a 29% year-over-year increase.

24

25   [46] *Biggest Companies in the World by Market Cap*, Investopedia (Oct. 16, 2024)
     https://www.investopedia.com/biggest-companies-in-the-world-by-market-cap-5212784 (last accessed

26   Nov. 25, 2024).

27   [47] *Alphabet Inc. Valuation Measures, Yahoo Finance*, https://finance.yahoo.com/quote/GOOG/key-
     statistics/ (last accessed Nov. 25, 2024).

28

188.   Alphabet's vicarious liability is further supported by: a) active participation in strategic decisions regarding AI development; b) knowledge of copyright implications through sophisticated legal resources; c) deliberate structuring of operations to capture value from infringing activities; d) failure to implement copyright compliance measures despite supervisory authority; and e) ongoing benefit from continued exploitation of infringing models.

189.   Alphabet's supervision of Google extends beyond mere corporate ownership: a) direct involvement in operational decisions affecting AI development; b) strategic oversight of AI integration across product lines; c) resource allocation for AI initiatives; d) coordination of cross-entity AI development teams; and e) management of corporate restructuring fundamental to AI development.

190.   The commercial impact of the supervised infringement includes: a) lost licensing revenue that would have been paid to copyright holders; b) diminished market value of copyrighted works; c) unfair competitive advantage in AI development; d) ongoing monetization of infringing models through product integration; and e) enhanced market position secured through accelerated development.

191.   Alphabet further failed to stop Google LLC's infringing activity. As set forth above, it not only financially benefitted directly from Google LLC's infringing activity, it took part in the infringing activity itself through its direct involvement in operational decisions affecting AI development at Google LLC; strategic oversight of AI integration across product lines, including those at Google LLC; resource allocation for AI initiatives at and within Google LLC, coordinating cross-entity AI development teams; and managing corporate restructuring fundamental to AI development.

192.   Through its supervisory role and direct financial interest, Alphabet is vicariously liable for Google's copyright infringement. Pursuant to 17 U.S.C. § 504, Plaintiffs and Class Members are entitled to: a) maximum statutory damages of $150,000 per work for willful infringement; b) actual damages and disgorgement of profits attributable to infringement; c) injunctive relief under 17 U.S.C. § 502; and d) recovery of full costs and reasonable attorney's fees under 17 U.S.C. § 505.

193.   Specifically, Alphabet's material contribution to the infringement manifests in: a) systematic facilitation of unauthorized reproduction and distribution of copyrighted works, undermining Plaintiff's exclusive rights under 17 U.S.C. § 106; b) knowing provision of essential means and

SECOND AMENDED CONSOLIDATED CLASS ACTION COMPLAINT

1   ~~infrastructure that enable large-scale appropriation of protected creative expression; c) continued support~~

2   ~~of comprehensive copying that exceeds any legitimate purpose or justification; and d) active enablement~~

3   ~~of market usurpation that directly displaces Plaintiff's existing and potential revenue streams.~~

4   ~~194.   The resulting harm is irreparable in nature. Through its supervision and control over~~

5   ~~Google's technical infrastructure and operations, combined with its direct financial interest in the~~

6   ~~infringing activities, Alphabet knowingly enables and substantially participates in conduct causing~~

7   ~~immediate and irreversible injury to Plaintiffs' rights and market position.~~

8   ~~195.   Unless enjoined by this Court, Alphabet will continue to benefit from and facilitate~~

9   ~~Google's ongoing infringement through continued operation and monetization of its AI models, causing~~

10   ~~irreparable injury for which Plaintiffs have no adequate remedy at law.~~

11   **VII.   PRAYER FOR RELIEF**

12   ~~196.~~182.   Wherefore, Plaintiffs request that the Court enter judgment on their behalf and on

13   behalf of the Class by ordering:

14   a) This action may proceed as a class action, with Plaintiffs serving as Class

15   Representatives, and with Plaintiffs' counsel as Class Counsel;

16   b) Judgment in favor of Plaintiffs and the Class against ~~Defendants~~Google;

17   c) An award of actual damages suffered by Plaintiffs as a result of the infringement and any

18   profits of the infringer that are attributable to the infringement, pursuant to 17 U.S.C.

19   § 504 (b);

20   d) An award of statutory damages, pursuant to 17 U.S.C. § 504 (c);

21   e) Destruction or other reasonable disposition of all copies ~~Defendants~~Google made or used

22   in violation of the exclusive rights of Plaintiffs and the Class, under 17 U.S.C. § 503(b);

23   f) Pre- and post-judgment interest on the damages awarded to Plaintiffs and the Class, and

24   that such interest be awarded at the highest legal rate from and after the date this class

25   action complaint is first served on ~~Defendants~~Google;

26

27

28

g) ~~Defendants are~~Google is to be ~~jointly and severally~~ responsible financially for the costs and expenses of a Court--approved notice program through post and media designed to give immediate notification to the Class;

h) Award injunctive relief as detailed above, and all other appropriate injunctive and equitable relief deemed just and proper;

i) Plaintiffs and the Class are awarded their reasonable costs and expenses incurred in this Action, including attorneys' fees, costs, and expenses; and/or

j) Further relief for Plaintiffs and the Class as the Court deems appropriate.

VIII.   JURY TRIAL DEMANDED

Under Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury of all the claims asserted in this Complaint so triable.

Dated: ~~December 20, 2024~~September 25, 2025   Respectfully submitted,

By: */s/ Joseph R. Saveri*
Joseph R. Saveri (SBN 130064)
Christopher K.L. Young (SBN 318371)
Elissa A. Buchanan (SBN 249996)
Evan Creutz (SBN 349728)
Aaron Cera (SBN 351163)
Louis Kessler (SBN 243703)
Alex Zeng (SBN 360220)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1505
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
jsaveri@saverilawfirm.com
cyoung@saverilawfirm.com
eabuchanan@saverilawfirm.com
ecreutz@saverilawfirm.com
acera@saverilawfirm.com
lkessler@saverilawfirm.com
azeng@saverilawfirm.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027

Telephone: (323) 968-2632
Facsimile: (415) 395-9940
mb@buttericklaw.com

Brian D. Clark (admitted *pro hac vice*)
Laura M. Matson (admitted *pro hac vice*)
Arielle Wagner (admitted *pro hac vice*)
~~Eura Chang~~Consuela Abotsi-Kowu (admitted
pro hac vice)
LOCKRIDGE GRINDAL NAUEN PLLP
100 Washington Avenue South, Suite 2200
Minneapolis, MN 55401
Telephone: (612)339-6900
Facsimile: (612)339-0981
bdclark@locklaw.com
lmmatson@locklaw.com
aswagner@locklaw.com
~~echang~~cmabotsi-kowu@locklaw.com

Stephen J. Teti (admitted *pro hac vice*)
LOCKRIDGE GRINDAL NAUEN PLLP
265 Franklin Street, Suite 1702
Boston, MA 02110
Telephone: (617) 456-7701
sjteti@locklaw.com

*Attorneys for Plaintiffs and the Proposed
Class in the* Zhang *Action*

Dated: ~~December 20, 2024~~September 25, 2025

/s/ *Ryan J. Clarkson*
Ryan J. Clarkson (SBN 257074)
Yana Hart (SBN 306499)
Mark Richards (SBN 321252)
Tiara Avaness (SBN 343928)
CLARKSON LAW FIRM, P.C.
22525 Pacific Coast Highway
Malibu, CA 90265
Telephone: 213-788-4050
rclarkson@clarksonlawfirm.com
yhart@clarksonlawfirm.com
mrichards@clarksonlawfirm.com
tavaness@clarksonlawfirm.com

Tracey Cowan (SBN 250053)
CLARKSON LAW FIRM, P.C.
95 Third Street, Second Floor
San Francisco, CA 94103
Tel. (213) 788-4050
tcowan@clarksonlawfirm.com

Dated: September 25, 2025

/s/ *Lesley E. Weaver*
Lesley E. Weaver (SBN 191305)
Anne K. Davis (SBN 267090)
Joshua D. Samra (SBN 313050)
BLEICHMAR FONTI & AULD LLP
1330 Broadway, Suite 630
Oakland, CA 94612
Tel. (415) 797-2617
lweaver@bfalaw.com
adavis@bfalaw.com
jsamra@bfalaw.com

Gregory S. Mullens (admitted *pro hac vice*)
BLEICHMAR FONTI & AULD LLP
75 Virginia Road, 2nd Floor
White Plains, New York 10603
Telephone: (415) 445-4006
gmullens@bfalaw.com

Attorneys for Plaintiffs and the Proposed Class in the ~~in in the~~ *Leovy Action*